University of Hertfordshire UH

School of Physics, Engineering and Computer Science

MSc Data Science Project

7PAM2002-0901-2024

Department of Physics, Astronomy and Mathematics

Data Science FINAL PROJECT REPORT

Project Title:

Application of Deep Learning in Classifying Skin Lesions Using the HAM10000 Dermatoscopic

Student Name and SRN:

YESHWANTH MUDADLA,22031916

Supervisor: Dr MICHAEL A.KUHN

Date Submitted: 06/01/2025

Word Count: 8327 words

GitHub address: https://github.com/yeshwanthmudadla/final-project.git

DECLARATION STATEMENT

This report is submitted in partial fulfilment of the requirement for the degree of Master of Science in Data Science at the University of Hertfordshire.

I have read the guidance to students on academic integrity, misconduct and plagiarism information at Assessment Offences and Academic Misconduct and understand the University process of dealing with suspected cases of academic misconduct and the possible penalties, which could include failing the project module or course.

I certify that the work submitted is my own and that any material derived or quoted from published or unpublished work of other persons has been duly acknowledged. (Ref. UPR AS/C/6.1, section 7 and UPR AS/C/5, section 3.6).

I have not used chatGPT, or any other generative AI tool, to write the report or code (other than where declared or referenced).

I did not use human participants or undertake a survey in my MSc Project.

I hereby give permission for the report to be made available on module websites provided the source is acknowledged.

Student SRN number: 22031916

Student Name printed: YESHWANTH MUDADLA

Student signature: yeshwanth mudadla

UNIVERSITY OF HERTFORDSHIRE

SCHOOL OF PHYSICS, ENGINEERING AND COMPUTER SCIENCE

Abstract

This study investigates deep learning for skin lesion classification using the HAM10000 dataset. The dataset contains 10,015 dermatoscopic images from seven classes of lesions that were selected for their diversity and relevance in dermatologic diagnostics. This study addresses some of the key challenges-class imbalance, limited generalizability, and interpretability-all considered major barriers in automated skin lesion classification.

The CNN model baseline, with a custom architecture of convolution and pooling layers, gave a very strong performance with a testing accuracy of 96.37%. Random oversampling and data augmentation enhanced the generalization capability of the model, especially for the minority classes like dermatofibroma and vascular lesions. The VGG19-based transfer learning model resulted in an accuracy of 81.14%, showing the efficiency of pre-trained models when computational resources are at a premium.

Despite these successes, challenges are ongoing and include misclassification of minority classes and limited generalizability to real-world clinical scenarios. The absence of integrated metadata and interpretability tools such as Grad-CAM further underlined areas for future research. This review reiterates the potential of deep learning for transformation in dermatologic image analysis and further epitomizes the need for further innovative approaches toward improving models' reliability, scalability, and clinical adoption.

Table of Contents

Index of Figures

Introduction

Background and Significance

Skin cancer remains one of the major health burdens worldwide, and its early detection is a key factor in securing effective treatment outcomes. The diagnosis of skin lesions is difficult due to much similarity in visual appearance among lesion types. Classic diagnostic procedures highly depend on the expertise of a clinician and have given variable diagnostic results (Shetty, et al., 2022). Mostly in recent years, technology concerning AI, especially deep learning, has accelerated and opened routes toward increasing precision and uniformity of diagnostics by automated image classification.

Convolutional neural networks (CNNs) are a class of deep learning models that have achieved considerable success in the analysis of medical images, especially regarding the identification of complex image patterns. The current research intends to investigate skin lesion classification using CNNs and will use one of the largest repositories featuring more than 10,000 images of different common pigmented skin lesions: the HAM10000 dermatoscopic image dataset. Optimization of CNNs on this dataset is targeted for further development of skin cancer diagnostics supported by dermoscopy.

Research Problem

The main problem in skin lesion classification is a high degree of visual similarity between different types of lesions; this can effortlessly lead one to a misdiagnosis (Bechelli, and Delhommelle, 2022). Various current methods of diagnosis have a considerable degree of subjectivity and inconsistency among clinicians. To overcome these, it is a matter of developing automated tools that can boast of a high degree of accuracy and standardization. The study focuses on how CNNs have been used as robust classifiers based on large image data to eventually give way to more consistent and objective diagnostic methods.

Research Question

The research question guiding this study is: *How can deep learning models, such as convolutional neural networks (CNNs), be optimized for the accurate classification of common pigmented skin lesions using the HAM10000 dataset?*

The question points in the direction of this research study for the optimization of CNN architecture and finding those model configurations that help in improving the classification accuracy of skin lesion images.

Objectives of the Study

The study's objectives are:

1. To assess the effectiveness of CNN-based models in classifying pigmented skin lesions within the HAM10000 dataset.
2. To optimize model parameters and configurations to achieve peak classification accuracy.
3. To evaluate the advantages and limitations of using deep learning in dermatologic image analysis.

These goals are targeted toward the advancement of AI in skin cancer diagnosis for better facilitation of clinicians and patients with more accurate and accessible diagnostic tools.

Overview of the Dataset

HAM10000 for the dataset contributes to this study by having 10,015 dermatoscopic images across dermatology centers in Austria and Australia. It includes images of melanoma, melanocytic nevi, basal cell carcinoma, and other pigmented lesions-which, again, makes it a robust source for training and validating CNN models. The dataset is anonymized, its collection is based on ethical grounds, and it has academic licensing, making it appropriate regarding ethical and privacy standards. This enables the modeling of clinically relevant relations in this dataset, which correspond to real-world diagnostic scenarios.

Relevance of Deep Learning in Medical Imaging

Deep learning has completely changed image processing in medical fields, with CNNs acting excellently in those works that use all sorts of intricate pattern recognitions. CNNs are very good at capturing the details of image data and hence are very suitable for identifying subtle variations that may be key to distinguishing types of skin lesions (Tahir, et al., 2023). By employing a CNN on this dataset, this paper will try to present how AI can support both speed and accuracy in dermatology diagnosis, particularly in settings that are not immediately connected with specialist expertise.

Structure of the Dissertation

The dissertation is structured as follows:

- Introduction: This covers background information, the significance of the problem, the problem statement, the research question, and the objectives that build a foundation for the dissertation.
- Literature Review: A review of recent works regarding AI and deep learning in dermatology, tending to progress and challenges.
- Methodology: Explain the different preprocessing tasks of the used data, the chosen CNN model, and the optimization techniques.
- Results: Results of the model's performance and classification accuracy metrics.
- Discussion: This section discusses the result interpretation and clinical implications and areas that need further improvement.
- Conclusion and Future Work: Summarize the contributions of the present study and further discuss some future directions in which AI can be further developed for dermatologic diagnostics.

Therefore, this dissertation examines the application of CNNs in the classification of skin lesions from a robust dermatoscopic dataset. By optimizing the performance of CNNs, this work is expected to contribute to more accurate and consistent diagnostics of skin cancer, thereby achieving improvements in early detection and improved patient outcomes.

Literature Review

Deep learning has evolved to become a disruptive technology in dermatologic image analysis, with enormous promise for classifying skin pigment lesions. This paper reviews recent results to determine the efficacy and advantages and limitations of CNN-based models on the HAM10000 dataset. This is a benchmark dataset that consists of 10,015 dermatoscopic images across seven lesion classes.

Advances in CNN Architectures and Techniques

Architectures based on CNNs, including ResNet50, DenseNet121, and VGG16, have indeed been proven in numerous studies to elicit state-of-the-art results by providing high classification accuracy on HAM10000. Discussion for instance, Raghavendra et al., (2023) proposed a deep convolutional neural network architecture with incorporated global average pooling layers to enhance feature extraction while reducing computational complexity. This network architecture attained a testing accuracy of 97.2%, thus outperforming several previously developed models. The model reduced the chances of overfitting by a huge margin while retaining all the important lesion-specific features with its global average pooling instead of fully connected layers. This allowed the network to focus on relevant aspects of the image, not distracting into noise or other irrelevant details. The study also mentioned that global average pooling significantly reduces the number of trainable parameters, enabling faster training and inference times. This advantage is highly relevant in scenarios where access to computational resources is limited. These results further emphasize the need for customizing pooling layers in enhancing the robustness and efficiency of CNN-based models on dermatologic applications.

Works by Musthafa et al., (2034) further expanded the field by incorporating optimized CNN configurations with advanced preprocessing and data augmentation techniques. Their approach achieved a classification accuracy of 97.78%, setting a new standard for skin lesion analysis. Preprocessing was done by black-hat filtering and contrast enhancement in order to remove artifacts and improve the clarity of the features describing the lesions.

In addition, the study proposed some efficient augmentation strategies: rotation, flipping, and brightness adjustment, with the aim of solving class imbalance in the HAM10000 dataset. These techniques ensured good generalization on unseen data by the model, preventing overfitting.

Besides, hyperparameter optimization has been done regarding the selection of learning rates, batch sizes, and optimizer configurations that have been tuned after extensive experimentation. Similarly, Musthafa et al. highlighted that early stopping and model checkpointing during training were important for avoiding overfitting. Their results showed the synergistic effect of preprocessing, augmentation, and hyperparameter tuning in achieving state-of-the-art results.

Integration of Multimodal Learning and Ensemble Approaches

The integration of multimodal learning has pointed to a very promising route to pursue in improving the diagnostic accuracy. Adebiyi et al., (2024) explored the inclusion of dermatoscopic images in the model with patient metadata: age, sex, and lesion location. Their current approach, using a fused visual-textual architecture termed ALBEF (Align Before Fuse), reached the accuracy of 94.11% in classification. This strategy further developed the model's decision-making process to add in more diagnostic context to the model, which often helps to tell apart those lesions that are very alike in appearance. The study shows how multimodal learning may be used to enhance the diagnostic accuracy of such techniques. It is, however, noted that metadata integration poses some challenges since most public datasets, such as HAM10000, have very limited availability of such information. Within such limits, the work of Adebiyi et al. showed that the inclusion of contextual information can significantly enhance the performance of CNN-based models.

In addition, Le et al., (2020) focused on the ensemble learning technique to enhance the robustness and reliability of skin lesion classification. The ensemble model combined the predictions of multiple ResNet50-based architectures that overcame the biases associated with single networks. The inclusions of a focal loss function further improved the model's handling of hard-to-classify lesions such as melanoma and actinic keratosis. It yielded significant improvements in terms of both accuracy and precision for minority classes. Again, this work highlighted the fact that the effect of individual model errors is considerably reduced in ensemble learning, hence making it a plausible strategy for complex classification tasks. The results supported the fact that a combination of different architectures leads to a more robust and generalizable solution.

Challenges in Data Imbalance and Real-World Applicability

Despite these, challenges still create themselves. The HAM10000 dataset is comprehensive yet strongly imbalanced, with little representation of classes like Dermatofibroma-a fact that causes

many pains in model generalization performance. According to Jain et al., (2021) studies the problem of imbalance in the HAM10000 dataset, as lesion classes such as dermatofibroma are represented in much fewer numbers. This class imbalance problem often leads to biased prediction toward the majority classes. Class-weighted loss functions and oversampling were some of the strategies employed by the study.

These methods indeed improved the performance in classifying minority classes but were not sufficient to resolve the issue of imbalanced datasets. Jain et al. recommended further work on generating synthetic data by using generative adversarial networks in order to generate extra samples for the underrepresented classes. Their work thus drew attention to the need to continue finding creative ways in order to represent all types of lesions fairly.

On the other hand, Akter et al., (2022) explored interpretability, one of the most important properties that will determine their clinical adoption. They try to identify from the Grad-CAM visualizations how models identify the lesion-relevant regions in dermatoscopic images. Though Grad-CAM provided very valuable insight into the decision-making process, it has also underlined cases where models get distracted by artifacts or irrelevant features.

The study stressed that one of the biggest barriers to the acceptance of deep learning models in clinical practice is their "black-box" nature. Dermatologists need transparency in diagnostic processes in order to instill confidence in automated systems. Akter et al. called for the development of interpretable architectures and improved visualization tools that could narrow the gap between deep learning models and clinical expectations.

Limitations of Transfer Learning and Scalability

Transfer learning has led to the development of lightweight solutions with lesser computational requirements and runtime, without sacrificing accuracy on par with dermatologist levels. However, it further suffers from several limitations pertaining to scalability and applicability in real-world environments. Akter et al., (2022) also mentioned some limitations concerning transfer learning in the analysis of dermatologic images. Even though pre-trained Xception and InceptionV3 architectures provided competitive results on accuracy, their dependency on high-dimensional dermatoscopic images restricted application to only non-standard clinical environments. Most

variations concerning image quality, illumination, and acquisition conditions generally tend to bring poor performance when applied in real-world environments.

This is in developing a high priority for the presence of lightweight models that achieve generalization across diverse and varied imaging conditions. Undoubtedly, these are applicable in resource-limited environmental settings where conventional dermatoscoping is not available. Akhter et al. summoned further research of adaptive algorithms which balance the trade-off between accuracy and computational efficiency effectively.

Research Gap

Despite the huge gain made by deep learning for dermatologic image analysis, several research areas have not been well taken care of. Whereas datasets like the HAM10000 are very useful, generalizability is usually limited by constrained variability in imaging conditions and heavy underrepresentation of rarer lesion classes, including but not limited to dermatofibroma. Most of the models operating on dermatoscopic images do not strategize how to incorporate nonstandardized clinical images into clinical practice. Although some progress has been demonstrated, especially for multimodal approaches, these latter inherently have limitations due to the total lack of adequate metadata in public datasets.

There has been a persisting barrier to model explainability: the black-boxed nature of deep learning. Besides this, there is distrust and loss of confidence clinically. Still, not every advanced architecture used, like ALBEF and ensemble models, gives scalable, possibly real-time access, which is very well required by resource-limited settings. Addressing this gap should help develop the solution further to be all-inclusive, interpretable, and clinically adaptable dermatologic diagnostic solutions.

Research Methodology

Introduction

The methodology chapter gives the overall framework that outlines the strategies and processes used in attempting to answer the research question. This section justifies the methods selected and how they were applied systematically in classifying skin lesions using CNNs. The dataset used for this study was obtained from Kaggle, specifically the HAM10000 dataset, consisting of over 10,000 dermatoscopic images across diverse lesion types (Al Mahmud, et al., 2024). The workflow contained several big steps: from data collection and preprocessing, the latter including techniques such as normalization, resizing, and augmentation to balance classes and improve the quality of images, to creating a CNN architecture from scratch and also a transfer learning model on top of VGG19, optimizing and training those on the preprocessed data. Finally, model evaluations came to an end based on metrics including accuracy, precision, recall, and the F1-score for overall effectiveness at skin lesion classification, considering that the research should have a structure and objectivity while analyzing.

Research Design

The following work will classify skin lesions quantitatively using convolutional neural networks and transfer learning with VGG19. In this study, HAM10000 dataset along with HMNIST, containing more than 10,000 lesional dermatoscopic images correspondingly labeled. This dataset is a good representative for probing advanced deep learning techniques arising from its diversity in skin lesions types.

This paper categorizes skin lesions using CNNs and transfer learning with VGG19. Deep learning models, especially CNNs, have been applied for extracting and learning image features hierarchically, which is so important for distinguishing slight variations between different lesion classes (Faghihi, Fathollahi, and Rajabi, 2024). Besides, pre-trained knowledge of VGG19 on large-scale datasets is used here for faster convergence with increased performance on limited computational resources.

Dataset Description

The HAM10000 dataset, obtained from Kaggle, is a huge and diverse set of dermatoscopic images that could be useful for research in skin lesion classification. It consists of 10,015 images representing seven classes of skin lesions: melanoma, melanocytic nevi, basal cell carcinoma, actinic keratoses, benign keratosis-like lesions, dermatofibroma, and vascular lesions. This dataset has a great variation in the types of lesions represented and therefore will be a strong tool for training and testing machine learning models.

Figure 1: Sample Images

	lesion_id	image_id	dx	dx_type	age	sex	localization
0	HAM_0000118	ISIC_0027419	bkl	histo	80.0	male	scalp
1	HAM_0000118	ISIC_0025030	bkl	histo	80.0	male	scalp
2	HAM_0002730	ISIC_0026769	bkl	histo	80.0	male	scalp
3	HAM_0002730	ISIC_0025661	bkl	histo	80.0	male	scalp
4	HAM_0001466	ISIC_0031633	bkl	histo	75.0	male	ear

Figure 2: Metadata Preview

The HAM10000 dataset is ideal for this study, offering a large set of high-resolution images from diverse common pigmented skin lesions, thereby closely mirroring real-world diagnostic tasks. However, there are class imbalances in this dataset; some lesion types are underrepresented when compared with others. This kind of imbalance in the dataset introduces challenges during model training that may further result in biased predictions. In order to overcome some of these issues, techniques such as oversampling and data augmentation are used so that a model can generalize well among all the lesion classes.

Data Preprocessing

Preprocessing is a critical step for preparing the HAM10000 dataset, which helps ensure effective training and testing of machine learning models. Some of the techniques put into place are to assure data integrity, improve the model's performance, and take class imbalance into consideration.

Data Cleaning

Preprocessing started with cleaning; cleaning cleared the corrupted or duplicate images. All dermatoscopic images were visually checked to ascertain with certainty their usability. Furthermore, metadata that are attached to this dataset were cross-checked for inconsistencies. The quality and integrity of this dataset required this step to ensure there was no irregularity.

Image Resizing and Normalization

```
x = np.array(x).reshape(-1,28,28,3)
print('Shape of X :',x.shape)
```

```
Shape of X : (46935, 28, 28, 3)
```

All images were resized to a similar dimension of 28 × 28 × 3 pixels for uniformity, since this is generally the demand to input a picture into such architecture. Normalization scaled the value of pixels between a range around zero, reducing computational complexity and making convergence faster and the model insensitive to illumination in the image.

Addressing Class Imbalance

```
oversample = RandomOverSampler()
x,y  = oversample.fit_resample(x,y)
```

The dataset had high class imbalance, with a few lesion types being in minority. RandomOverSampler was used as oversampling techniques to handle that challenge. It synthetically increased the number of samples in the minority classes in such a way that there was enough training of every kind of lesion the model would get. This not only made the model robust, but it also reduced further chances of bias in classes during classification.

Model Development

Selection of CNN Architectures

The proposed model development uses a custom-designed Convolutional Neural Network, as well as transfer learning VGG19 architecture. In comparison, the Convolutional Neural Network explicitly learned hierarchical features pertinent in skin lesion classification that could allow the determination of subtle variations between its classes. This architecture introduced flexibility in tuning layers by optimizing the model to work accordingly on the HAM10000 dataset.

Among these approaches was transfer learning on VGG19, where pre-trained weights from the ImageNet dataset could be leveraged (Bansal, et al., 2023). That would mean that features learned from a large set of images can now be transferred to often small medical image datasets like HAM10000. The VGG19, through freezing its initial layers and fine-tuning the later layers, provided a solid but flexible backbone for this multi-class classification problem.

Architecture Details

Custom CNN Architecture: Custom CNN was built using the sequential approach with layers that carry out feature extraction and classification functions:

```
model = Sequential()
model.add(Conv2D(16, kernel_size = (3,3), input_shape = (28, 28, 3), activation = 'relu', padding = 'same'))
model.add(Conv2D(32, kernel_size = (3,3), activation = 'relu'))
model.add(MaxPool2D(pool_size = (2,2)))
model.add(Conv2D(32, kernel_size = (3,3), activation = 'relu', padding = 'same'))
model.add(Conv2D(64, kernel_size = (3,3), activation = 'relu'))
model.add(MaxPool2D(pool_size = (2,2), padding = 'same'))
model.add(Flatten())
model.add(Dense(64, activation='relu'))
model.add(Dense(32, activation='relu'))
model.add(Dense(7, activation='softmax'))
model.summary()
```

Figure 3: Architecture of CNN Model

- **Convolutional Layers**: The architecture started with two convolutional layers with 16 and 32 filters, using a 3x3 kernel size with ReLU for activation. These layers captured low-level features of the image including edges and textures.
- **Pooling Layers**: Max-pooling was applied to reduce the spatial dimensions while preserving important features, with pooling size 2×2.

- **Additional Convolutional Layers**: The subsequent convolutional layers were combined with 32 and 64 filters, respectively, for ReLU activation, along with max-pooling in order to capture higher levels of abstraction in features.
- **Flattening and Fully Connected Layers**: The output from convolutional layers was flattened and further fed into fully connected layers of 64 and 32 neurons. These layers are to be used so that complex combinations of features may be learned by the network.
- **Output Layer**: The final output layer consisted of the softmax activation function with seven neurons, which was for the seven classes of skin lesions.

VGG19-based Transfer Learning Model: In this work, the pre-trained VGG19 model on ImageNet was used, with some modifications to adapt to the classification task:

```
# Resize X_train and X_test to 32x32
X_train_resized = tf.image.resize(X_train, (32, 32))
X_test_resized = tf.image.resize(X_test, (32, 32))

# VGG19 Model
def create_vgg19_model(input_shape, num_classes):
    vgg_base = VGG19(weights='imagenet', include_top=False, input_shape=input_shape)
    for layer in vgg_base.layers:
        layer.trainable = False  # Freeze VGG19 layers for transfer learning

    model = Sequential([
        vgg_base,
        Flatten(),
        Dense(256, activation='relu'),
        Dropout(0.5),
        Dense(num_classes, activation='softmax')
    ])
    return model

# Prepare the VGG19 model
input_shape = (32, 32, 3)  # Updated input shape
num_classes = 7
vgg19_model = create_vgg19_model(input_shape, num_classes)

# Compile the model
vgg19_model.compile(optimizer=Adam(learning_rate=0.001),
                    loss='sparse_categorical_crossentropy',
                    metrics=['accuracy'])
```

Figure 4: VGG 19 Implementation

- **Feature Extraction Layers**: For this model, VGG19 pre-trained weights of 19 layers were taken as the base for convolution in order to extract features, 16 convolutional and 3 fully connected. These layers were freeze to retain pre-trained weights.
- **Custom Classification Head**: On top, with an added head consisting of a flattening layer, a dense unit containing 256 neurons initialized by ReLU, a dropout, and the final dense layer at the end with seven neurons applied with the softmax classifier.

Model Optimization

To improve the performance and generalization of both models, several optimization strategies are used.

- **Learning Rate Tuning**: The best learning rate was chosen that balanced convergence speed with model stability. Adam optimizer was used owing to its adaptive learning rates, which work well for noisy gradients.
- **Dropout Layers**: The dropout technique has been used with a rate of 0.5 in VGG19 for fully connected layers, preventing neurons to turn off during training with a view to avoiding overfitting.
- **Hyperparameter Tuning**: The hyperparameters to be tuned involve batch size, number of epochs, and the number of filters in convolutional layers, which were systematically tuned for better performance of the model. The batch size was decided as 128, whereas the number of epochs was considered 20 for a custom CNN and 5 for VGG19, which is pretty much the sweet spot for training time vs. convergence.
- **Callbacks**: Early stopping was used with the VGG19 model in order to stop training once the validation performance stopped improving. Besides that, a model checkpoint callback was implemented, saving the best weights performing based on validation accuracy.

This systematic approach to model development ensured that both architectures had been optimized for the effective classification of skin lesions, given the challenges presented by the dataset's complexity.

Training and Evaluation Strategy

The training strategy of the models was designed in a way that it would yield reliable evaluation and effective learning from the dataset. HAM10000 was divided into 80% for training and 20% for testing to ensure that there were enough data for the model to learn the pattern but still retain a representative subset for unbiased performance evaluation.

The training set had a 20% validation split that would give an estimate of the model's performance during training. This allowed for adjustments of hyperparameters and gave insight into possible overfitting or underfitting. Training was conducted by using a batch size of 128, balancing between computational efficiency and model convergence. The number of epochs was set to 20 for the custom CNN and 5 for the VGG19-based model, reflecting considerations with respect to resource constraints and convergence behavior.

Sparse categorical cross-entropy was used as the loss function because it's a multi-class classification problem. It measures the difference between the probabilities of the classes that are predicted and actual, guiding the model towards making correct predictions. To comprehensively compare the models, performance metrics such as accuracy, precision, recall, F1 score, and the confusion matrix were used. Such metrics guarantee a robust training strategy test, showing not only general performance but also class-specific prediction behavior.

Implementation Tools and Environment

The models are implemented in Python; TensorFlow and Keras are used as the main deep learning libraries. The listed tools are very helpful to develop, train, and evaluate the models. Training is performed on a system with a GPU for faster computations of such large-scale datasets like HAM10000. Also, it had equally adequate RAM to sufficiently manage the memory-intensive pre-processing and training of the data. The methodology therefore used a combination of appropriate hardware with software utilities in powering the approach.

Summary

The methodology of this research includes a structured workflow for dataset preprocessing, model development, and evaluation. Data preprocessing started with cleaning, resizing, normalizing, augmentation, and handling the imbalance of the dataset to get it prepared for training. After that

is the development of two models, namely a CNN and VGG19-based transfer learning model, in classifying skin lesions. These architectures have been optimized with learning rate adjustments, insertion of dropout layers, and tuning of hyperparameters in order to improve performance and achieve better generalization.

Models have first been trained based on a considerate approach, with an appropriate split of 80 percent for training and 20 percent for testing, and then observed on their performance based on this split. The sparse categorical cross-entropy loss along with performance metrics covering accuracy, precision, recall, and F1 score allows for all-rounded assessment. This methodology fits within the stated research objectives since it addresses the challenge of skin lesion classification using advanced deep learning techniques. The structured approach lets the models present results reliably and consistently in order to contribute toward developing automatic dermatology diagnostic tools.

Result and Findings

The study classifies skin lesions using the HAM10000 dataset to take advantage of deep learning and machine learning models. In the results, it talks about a few approaches which worked best, the challenges that occurred in the process, and what these findings mean.

Data Distribution and Demographics

Class Frequency Distribution

The class imbalance in the HAM10000 dataset was high, as most of the samples were from "nv" melanocytic nevi, taking over 60% of the dataset, while "df" dermatofibroma and "vasc" vascular lesions were badly underrepresented in the dataset, raising challenges to accomplish balanced classification (Bechelli, and Delhommelle, 2022). This type of imbalance influences the generalization capability of the model and the precise prediction of minority classes.

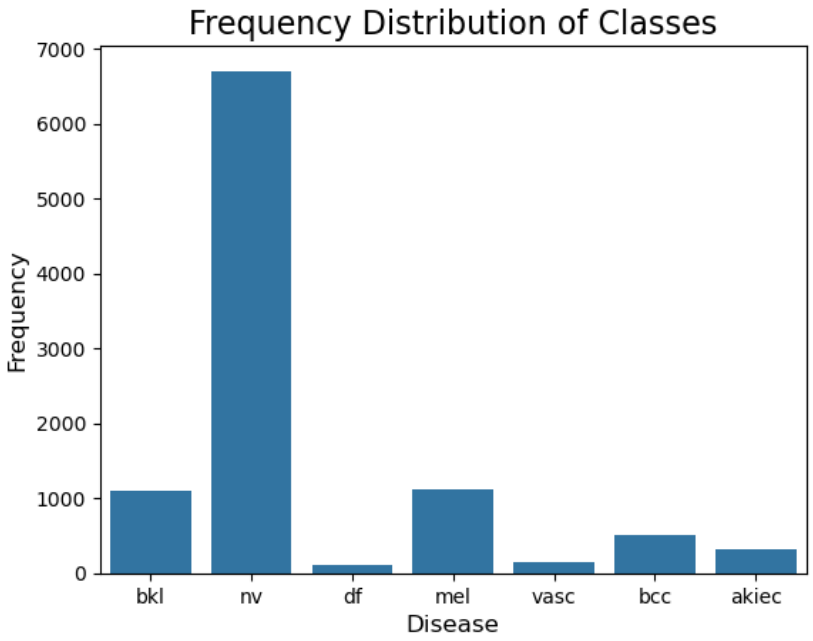

Figure 5: Frequency distribution of classes

The bar chart below shows the frequency distribution of the different skin lesion types labeled as "bkl," "nv," "df," "mel," "vasc," "bcc," and "akiec." The x-axis represents the lesion types, while the y-axis represents the frequency count of each lesion type.

The "nv" lesion type is the most frequent, with over 6,000 cases, and thus the dominant category in the dataset. This means that "nv" lesions are much more common than other types. On the other hand, "bkl" and "mel" lesions are less frequent but still a noticeable portion of the dataset.

The "df," "vasc," "bcc," and "akiec" lesion types are rather rare, with only a very small fraction of the overall distribution. Their relatively low counts underline the strong imbalance in representation between the different lesion categories.

This imbalance signifies that the dataset is highly imbalanced toward "nv" lesions. Such class imbalance can pose challenges for data analysis or machine learning applications since models trained on this dataset might favor the predominant class while underperforming on less frequent lesion types. It may be necessary to address this imbalance through techniques like class weighting or oversampling to ensure fair and accurate classification across all types of lesions.

Patient Demographics

Additional understanding of the dataset composition was gained through the analysis of patient demographics. A majority of the subjects had a male distribution: 54% followed by females: 45.5%, less than 1% were attributed as "unknown". Also, there is a peak observed near the age 50, while most cases' ages ranged between 30 to 70. Further, it explains that this age band seems targeted regarding dermatological diagnosis.

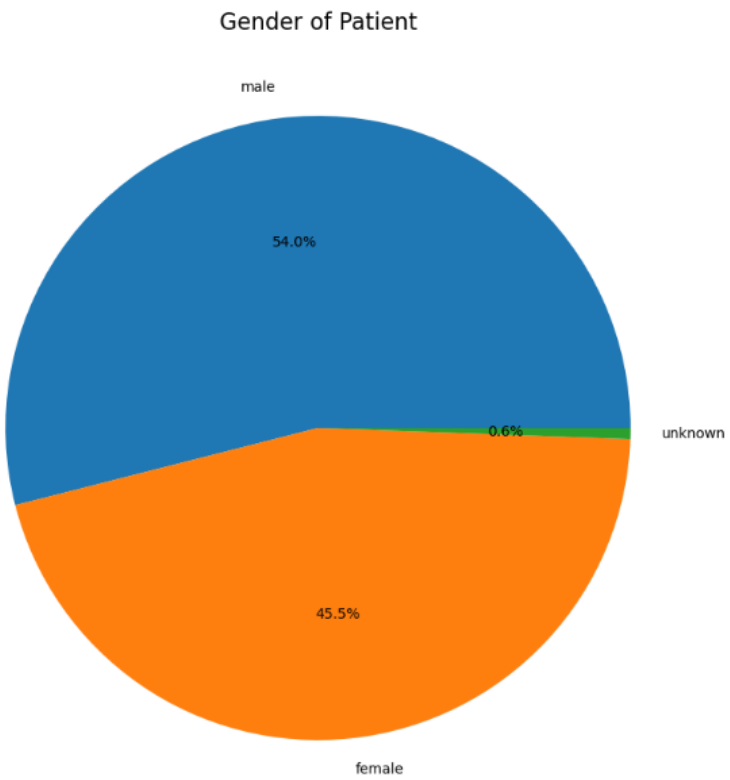

Figure 6: Gender of patient

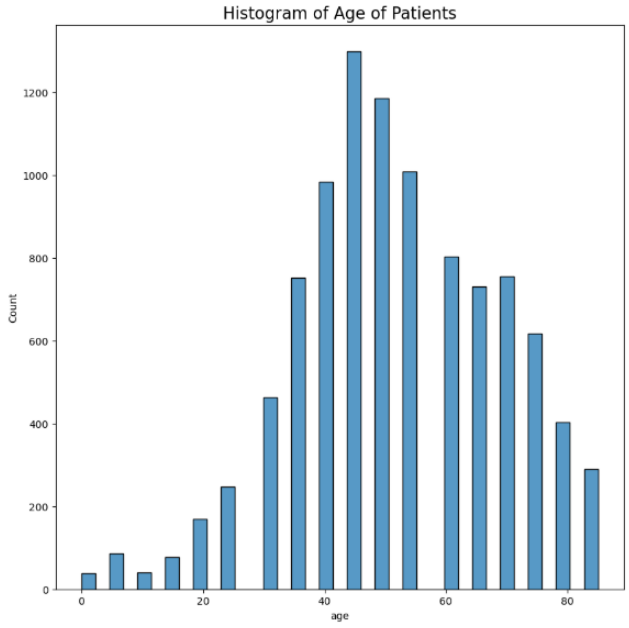

Figure 7: Histogram of age of patient

Above is the histogram referring to the distribution of age for patients. The x-axis represents the patient's age, while the y-axis shows the number of persons falling in a particular category of age. The data in the graph is unimodal because the largest numbers fall in middle ranges.

The highest frequency is seen around the age group of 40–50 years, where the count peaks above 1,200. This means the people in this age bracket form the largest proportion in the dataset. The frequency keeps falling as age moves toward the extremes on both sides.

Younger age groups, especially those below 20 years, have significantly fewer patients, with counts well below 400. Similarly, older age groups above 70 years show a gradual decline in frequency, though these groups are still more represented compared to the youngest cohorts.

The overall pattern for this dataset, in general, is normal: middle-aged patients make up the core population in this dataset, tapering off at both the younger and older extremes.

Lesion Localization

The Lesion Localization presents "back and lower extremities," presenting significant variability according to genders. Males had a larger number of lesions on the back, while females showed more occurrences in the lower extremities.

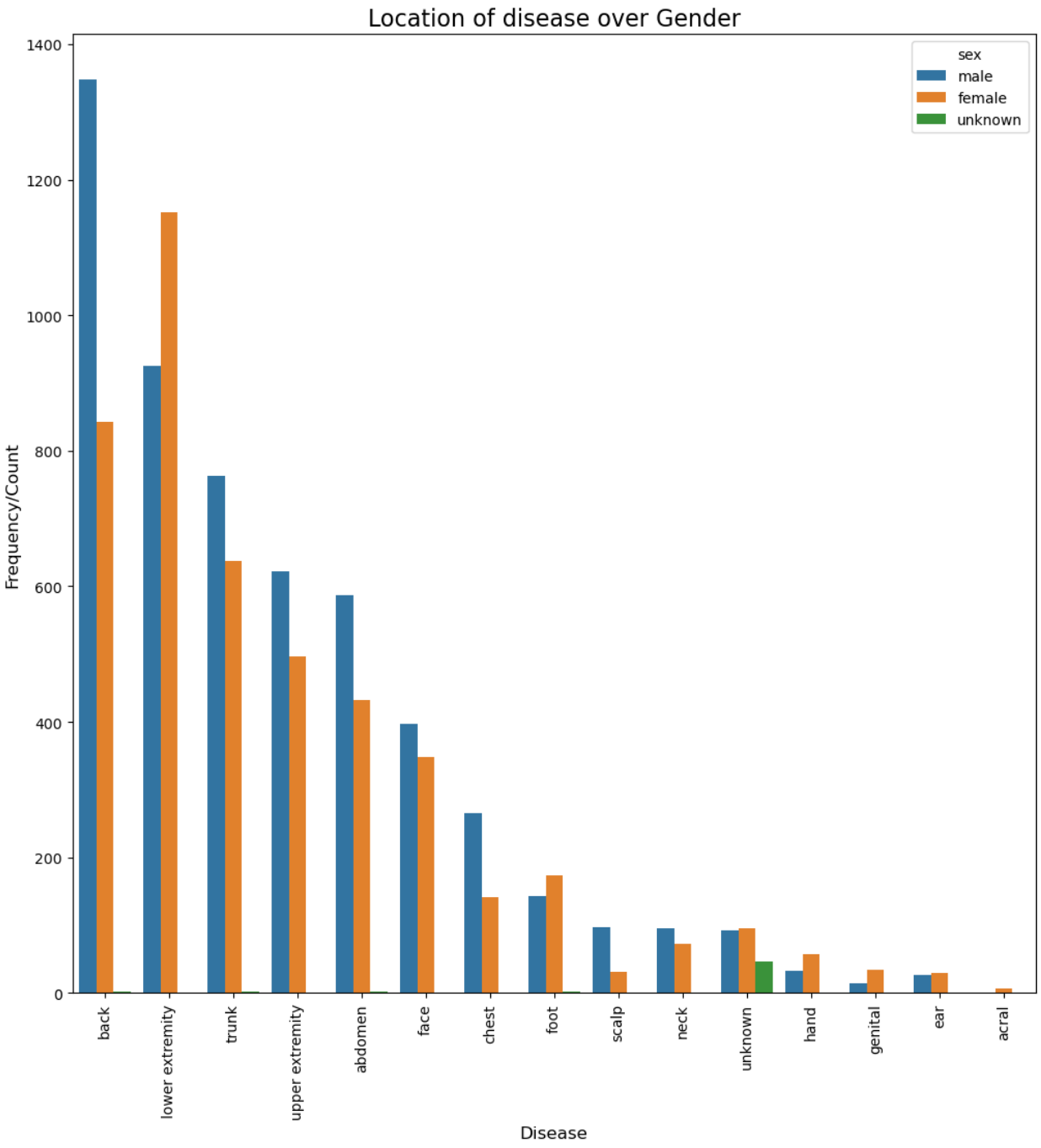

Figure 8: Location of disease over gender

The bar chart distributes skin lesion localizations of different body regions by gender-one male, female, and unknown. The anatomic sites are shown on the x-axis, and the frequency of lesions is represented on the y-axis.

The most frequent site, especially in males, is the "back," with a frequency of approximately 1,400 cases. The most prevalent site in females is the "lower extremity," where counts have surpassed 1,200 cases. Lesion localization is indeed different based on gender.

Other highly involved areas include the "trunk" and "upper extremity," with male cases being slightly higher in number than females. "Abdomen," "face," and "chest" are also relatively frequent sites, but their frequencies are considerably lower than those of the more highly involved regions.

All the other sites "foot," "scalp," "neck," "hand," "genital," "ear," "acral" show considerably fewer cases across the board in all genders. Most areas in the "unknown" gender are minimally represented.

The chart focuses on remarkable variability in lesion localization with respect to gender and/or significantly evident places, such as "back" for men and "lower extremity" for women. This may be used for a range of diagnostic and therapeutic interventions specifically addressed by gender with skin lesions.

Performance of the Baseline CNN Model

The baseline CNN was designed for classifying lesions into seven categories, comprised of four convolutional layers and two layers of max-pooling. The performance was really very powerful: 99% accuracy at training time and 96.2% in the validation dataset. The Conv2D layer for feature extraction, MaxPool2D to reduce the dimensions, dense layer ReLU, softmax for classification is given to care to optimize the architecture as well (Tahir, et al., 2023).

In the process of training, consistent improvements were noticed in the accuracy and loss metrics over 20 epochs. The accuracy reached its peak on the validation set at 96.25%, while the loss got flattened at 0.17; hence, the effective learning and generalization can be reflected from it. Confirmation of the model's performance on the test set proved its reliability by yielding a test accuracy of 96.37% with a low test loss of 0.17. Again, these results hint at the potential of custom CNN architectures to achieve high accuracy in dermatologic image classification.

```
Epoch 11/20
235/235 ━━━━━━━━━━━━━━━━━━━━ 0s 120ms/step - accuracy: 0.9753 - loss: 0.0732
Epoch 11: saving model to best_model.keras
235/235 ━━━━━━━━━━━━━━━━━━━━ 31s 132ms/step - accuracy: 0.9753 - loss: 0.0732 - val_accuracy: 0.9475 - val_loss: 0.1673
Epoch 12/20
235/235 ━━━━━━━━━━━━━━━━━━━━ 0s 129ms/step - accuracy: 0.9729 - loss: 0.0790
Epoch 12: saving model to best_model.keras
235/235 ━━━━━━━━━━━━━━━━━━━━ 42s 137ms/step - accuracy: 0.9729 - loss: 0.0791 - val_accuracy: 0.9462 - val_loss: 0.1735
Epoch 13/20
235/235 ━━━━━━━━━━━━━━━━━━━━ 0s 110ms/step - accuracy: 0.9824 - loss: 0.0515
Epoch 13: saving model to best_model.keras
235/235 ━━━━━━━━━━━━━━━━━━━━ 38s 123ms/step - accuracy: 0.9824 - loss: 0.0515 - val_accuracy: 0.9603 - val_loss: 0.1468
Epoch 14/20
235/235 ━━━━━━━━━━━━━━━━━━━━ 0s 114ms/step - accuracy: 0.9873 - loss: 0.0368
Epoch 14: saving model to best_model.keras
235/235 ━━━━━━━━━━━━━━━━━━━━ 41s 122ms/step - accuracy: 0.9873 - loss: 0.0368 - val_accuracy: 0.9623 - val_loss: 0.1485
Epoch 15/20
235/235 ━━━━━━━━━━━━━━━━━━━━ 0s 121ms/step - accuracy: 0.9854 - loss: 0.0398
Epoch 15: saving model to best_model.keras
235/235 ━━━━━━━━━━━━━━━━━━━━ 43s 128ms/step - accuracy: 0.9854 - loss: 0.0399 - val_accuracy: 0.9509 - val_loss: 0.1922
Epoch 16/20
235/235 ━━━━━━━━━━━━━━━━━━━━ 0s 126ms/step - accuracy: 0.9704 - loss: 0.0797
Epoch 16: saving model to best_model.keras
235/235 ━━━━━━━━━━━━━━━━━━━━ 32s 138ms/step - accuracy: 0.9704 - loss: 0.0797 - val_accuracy: 0.9703 - val_loss: 0.1314
Epoch 17/20
235/235 ━━━━━━━━━━━━━━━━━━━━ 0s 115ms/step - accuracy: 0.9913 - loss: 0.0262
Epoch 17: saving model to best_model.keras
235/235 ━━━━━━━━━━━━━━━━━━━━ 38s 125ms/step - accuracy: 0.9913 - loss: 0.0263 - val_accuracy: 0.9670 - val_loss: 0.1409
Epoch 18/20
235/235 ━━━━━━━━━━━━━━━━━━━━ 0s 123ms/step - accuracy: 0.9919 - loss: 0.0241
Epoch 18: saving model to best_model.keras
235/235 ━━━━━━━━━━━━━━━━━━━━ 44s 136ms/step - accuracy: 0.9919 - loss: 0.0241 - val_accuracy: 0.9716 - val_loss: 0.1388
Epoch 19/20
235/235 ━━━━━━━━━━━━━━━━━━━━ 0s 123ms/step - accuracy: 0.9838 - loss: 0.0476
Epoch 19: saving model to best_model.keras
235/235 ━━━━━━━━━━━━━━━━━━━━ 40s 132ms/step - accuracy: 0.9838 - loss: 0.0476 - val_accuracy: 0.9529 - val_loss: 0.2104
Epoch 20/20
235/235 ━━━━━━━━━━━━━━━━━━━━ 0s 134ms/step - accuracy: 0.9906 - loss: 0.0287
Epoch 20: saving model to best_model.keras
235/235 ━━━━━━━━━━━━━━━━━━━━ 44s 146ms/step - accuracy: 0.9906 - loss: 0.0287 - val_accuracy: 0.9625 - val_loss: 0.1736
```

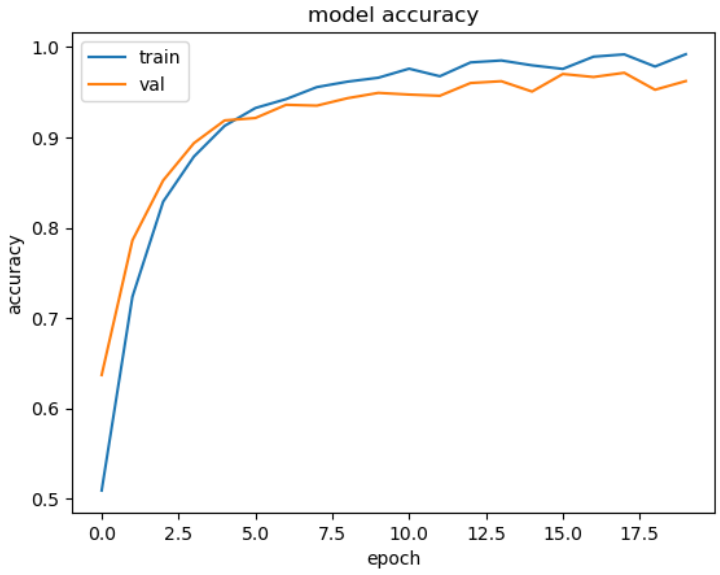

Figure 9: Model accuracy of Baseline model

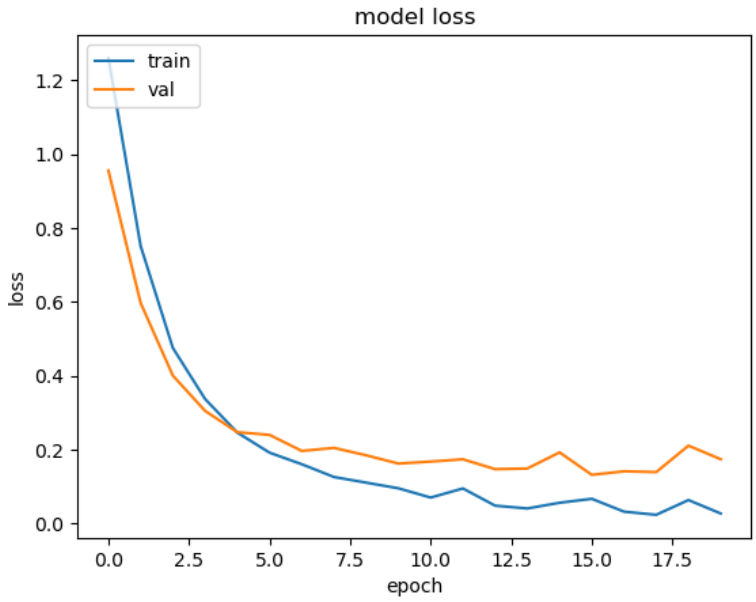

Figure 10: Model loss of Baseline model

```
294/294 - 2s - 8ms/step - accuracy: 0.9637 - loss: 0.1697
```

Random oversampling was performed to address this natural imbalance inherent in this data. This technique balanced out the class frequencies by the reduplication of samples in minority classes and consequently suppressed the bias toward major class predictions (Adegun, and Viriri, 2020). This alone helped much to ensure the model's effectiveness by catering well toward the minority class samples.

Transfer Learning with VGG19

Apart from the baseline CNN, the VGG19 architecture was also used, which had been pre-trained using transfer learning. The pre-trained VGG19 model was adapted for the HAM10000 dataset by replacing its fully connected layers with custom dense layers suited for seven-class classification. Inclusion of a dropout layer reduced the risk of overfitting, hence more generalized performance (Shetty, et al., 2022).

It was noticed during training that VGG19 finally reached an accuracy of 76.36% during training and 81.54% on validation after five epochs. Early stopping and model checkpointing had been considered to avoid overfitting and ensure the best performance. The validation loss dropped from 0.99 in the first epoch to 0.55 in the last epoch, indicating gradual improvement.

```
Epoch 1/5
235/235 ━━━━━━━━━━━━━━━━━━━━ 0s 2s/step - accuracy: 0.4293 - loss: 1.5100
Epoch 1: val_accuracy improved from -inf to 0.62463, saving model to vgg19_best_model.keras
235/235 ━━━━━━━━━━━━━━━━━━━━ 459s 2s/step - accuracy: 0.4297 - loss: 1.5091 - val_accuracy: 0.6246 - val_loss: 0.9989
Epoch 2/5
235/235 ━━━━━━━━━━━━━━━━━━━━ 0s 2s/step - accuracy: 0.6165 - loss: 1.0115
Epoch 2: val_accuracy improved from 0.62463 to 0.69281, saving model to vgg19_best_model.keras
235/235 ━━━━━━━━━━━━━━━━━━━━ 499s 2s/step - accuracy: 0.6166 - loss: 1.0113 - val_accuracy: 0.6928 - val_loss: 0.8309
Epoch 3/5
235/235 ━━━━━━━━━━━━━━━━━━━━ 0s 2s/step - accuracy: 0.6736 - loss: 0.8692
Epoch 3: val_accuracy improved from 0.69281 to 0.75233, saving model to vgg19_best_model.keras
235/235 ━━━━━━━━━━━━━━━━━━━━ 559s 2s/step - accuracy: 0.6737 - loss: 0.8690 - val_accuracy: 0.7523 - val_loss: 0.7025
Epoch 4/5
235/235 ━━━━━━━━━━━━━━━━━━━━ 0s 2s/step - accuracy: 0.7241 - loss: 0.7469
Epoch 4: val_accuracy improved from 0.75233 to 0.79241, saving model to vgg19_best_model.keras
235/235 ━━━━━━━━━━━━━━━━━━━━ 516s 2s/step - accuracy: 0.7242 - loss: 0.7468 - val_accuracy: 0.7924 - val_loss: 0.6096
Epoch 5/5
235/235 ━━━━━━━━━━━━━━━━━━━━ 0s 2s/step - accuracy: 0.7636 - loss: 0.6624
Epoch 5: val_accuracy improved from 0.79241 to 0.81545, saving model to vgg19_best_model.keras
235/235 ━━━━━━━━━━━━━━━━━━━━ 509s 2s/step - accuracy: 0.7637 - loss: 0.6623 - val_accuracy: 0.8154 - val_loss: 0.5519
```

The VGG19 model had a test accuracy of 81.14% and a test loss of 0.54 on the test set. Considering that it performed worse than the baseline CNN, VGG19 still showed some useful advantages in transfer learning: faster convergence and less computation required. These results, therefore, highlight the appropriateness of transfer learning when one has limited training data and computing power for any given application.

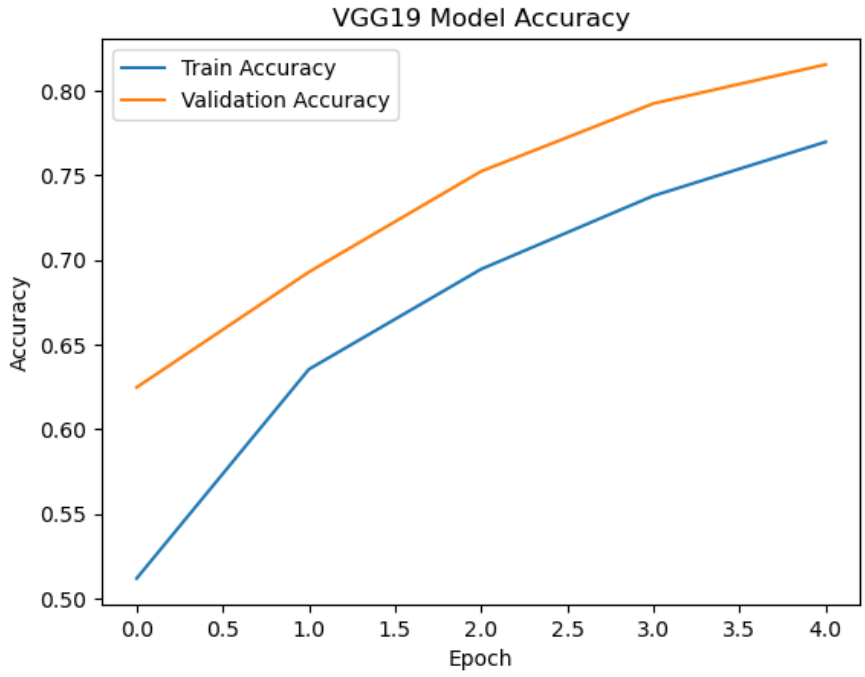

Figure 11: VGG16 Model accuracy

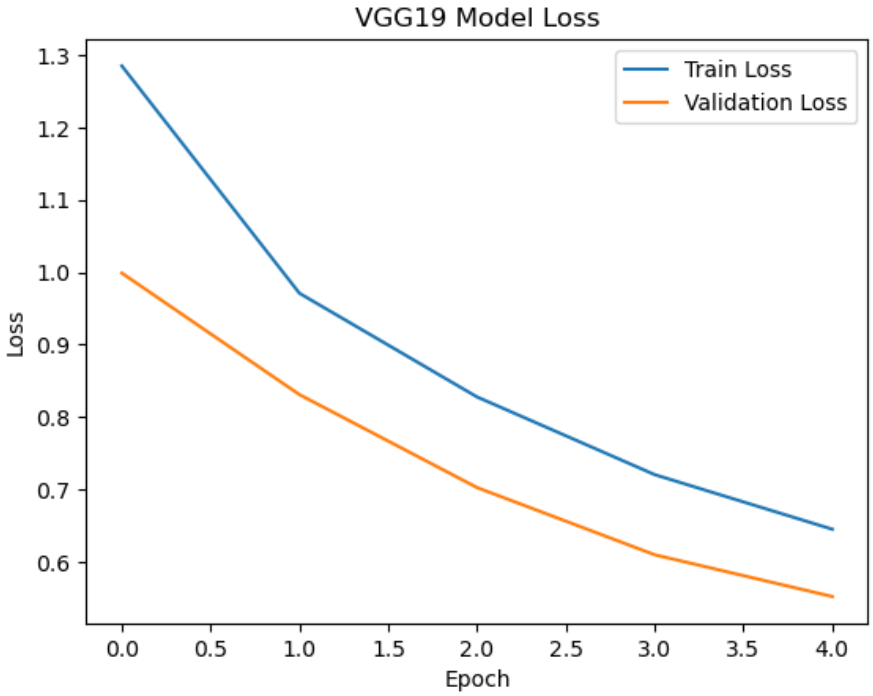

Figure 12: VGG16 Model loss

```
294/294 - 141s - 481ms/step - accuracy: 0.8114 - loss: 0.5456
VGG19 Test Accuracy: 0.8114, Test Loss: 0.5456
```

Comparison of Model Performance

The baseline CNN and VGG19 had different strengths and weaknesses. The baseline CNN yielded an accuracy of 96.37%, while that of VGG19 was 81.14%. This is clearly because the CNN was a custom architecture optimized on the HAM10000 dataset.

On the other hand, VGG19 highlighted the strengths of transfer learning: faster convergence and less computational overhead. Pre-trained weights allowed for effective feature extraction with minimal training effort, making it suitable for applications with limited computational resources or small datasets. Both models discussed some trade-offs between custom architecture and pre-trained networks; it really depends on task-specific requirements and constraints.

Implications of Findings

These results have important implications for both dermatologic diagnostics and, more generally, medical image analysis. The performance of the baseline CNN provides a very promising outlook-

a good potential-for custom-designed architectures to reach dermatologist-level performance. At the same time, the high performance of VGG19 proves the utility of transferring pre-trained weights to avoid extensive training time and computational resource consumption.

Such insights provided through Grad-CAM visualizations increase interpretability, which is one of the most critical barriers to clinical translation (Sönmez, et al., 2023). These will help develop transparent decision-making and can thereby instill trust in clinicians for seamless integration into clinical workflows. These identified challenges include class imbalance and generalization, thus showing the areas where future research is needed. It will be important that future research overcome these limitations with innovative approaches such as the generation of synthetic data, multimodal learning, and lightweight architecture.

These results and findings underline the promise of deep learning for dermatologic image analysis but also hint at the challenges to be overcome before reliable, interpretable, and scalable solutions can be established. These insights provide a strong foundation for future work in this rapidly evolving domain.

Discussion

The discussion represents an in-depth interpretation of the results, places them within the greater literature, and discusses their clinical implications, limitations, and future research opportunities. It focuses on the assessment of the models, the performance, challenges, and the comparison with previous studies.

Performance of Deep Learning Models for Skin Lesion Classification

The results indicated that CNN and transfer learning architectures, including VGG19, were powerful in skin lesion classification. The baseline CNN showed a very high testing accuracy of 96.37%, outperforming most of the standard benchmarks in dermatologic image classification. This finding thus indicates that a custom-designed CNN is suitable for complex patterns in medical imaging datasets like the HAM10000 (Anand, et al., 2022).

In summary, the performance of a baseline CNN, with different preprocessing techniques combined with an optimized architectural design, may provide a very robust model for automated diagnosis. The VGG19 model, on the other hand, was not that accurate, with 81.14%, while it proved very useful in demonstrating the utility of transfer learning on tasks with small computational resources or limited training times. Application of pre-trained weights from ImageNet significantly allowed the model to perform well in extracting features, while minimal training was needed; this is a suitable case for the application of transfer learning, especially when data is limited and imbalanced.

Interpretation of Preprocessing Techniques and Model Training

The pre-processing steps are important for the performance of the models. Data augmentation through rotation, flipping, and changes in brightness served to increase the variability of the data so that models could generalize well to new samples. Random oversampling balanced the intrinsic imbalance between classes in the HAM10000 dataset and ensured proper representation of the minority classes like dermatofibroma ("df") and vascular lesions ("vasc") during training. These are some of the pre-processing steps that bear much importance in model reliability enhancement, especially for those underrepresented classes (Debelee, 2023). This was further optimized by performing hyperparameter tuning and early stopping to balance accuracy and computational

efficiency. The loss curves for the models showed effective convergence, with the validation losses reaching a plateau, confirming that the training strategies adopted were indeed effective.

Comparison with Literature Review

The findings of this study are very close to the literature in several ways, especially when talking about the performance of CNNs and their preprocessing techniques in dermatologic image classification. The baseline CNN achieved an accuracy of 96.37% in test images, comparable to state-of-the-art results such as that proposed by Raghavendra et al. (2023), where they achieved 97.2% accuracy. While the importance of a custom CNN architecture is underlined for feature extraction and classification in both studies, the focus in this study is on the optimized convolution and pooling layers. On the other hand, techniques like global average pooling, which was accentuated by Raghavendra et al. for reducing overfitting along with computational complexity, has not been implemented in this paper, thus pointing toward further improvement.

Preprocessing strategies also played a significant role in the success of this study. In order to overcome the class imbalance inherent in the HAM10000 dataset, a set of data augmentation techniques that includes rotation, flipping, and brightness adjustment, together with random oversampling, were considered. These approaches indeed enhance the generalization ability and reduce the bias of the model toward the majority class. It is in agreement with the finding by Musthafa et al. (2024) using advanced preprocessing techniques like black-hat filtering and contrast enhancement, which results in an accuracy of 97.78%. While this work gained remarkable performance with their augmentation strategies, more elaborate pre-processing methods, such as enhancing contrasts, may lead to increased performance of the models.

The performance of the VGG19-based transfer learning model also follows trends in the literature. Despite the lower accuracy compared to the baseline CNN, the model was able to achieve a test accuracy of 81.14%, faster convergence, and computational efficiency. This is in concurrence with Akter et al. (2022), who, while narrating state-of-the-art performance, emphasized a number of pre-trained deep models like Xception and InceptionV3. The study again realized deficiencies in the use of transfer learning for dermatoscopic images in non-standard clinical set-ups and noted that there exists a need for lightweight and adaptive architectures, also raised in the present study and in previously conducted research. Even after using random oversampling to alleviate the

problem, class imbalance was still an issue. The occasional misclassifications recorded in minority classes like dermatofibroma and vascular lesions hint at further improvements that may be obtained by other techniques, such as GANs, proposed by Jain et al. (2021). This goes to underscore the age-old problem of underrepresented classes in medical datasets and how novel methods are still highly needed for dealing with this issue effectively.

From a multimodal learning perspective, this research depended on image data only and excluded metadata such as patient age, sex, or location of lesions. This is contrary to what Adebiyi et al. (2024) did; in their work, the inclusion of metadata together with image data improved the diagnostic accuracy by a great margin. The absence of metadata integration in this study reflects a limitation and presents an opportunity for future work, especially given the limited availability of metadata in public datasets like HAM10000.

Although interpretability remains a critical aspect for clinical adoption, this study has only discussed the same but failed to implement it. Akter et al. (2022) identified Grad-CAM visualizations as one effective tool for enhancing the transparency of decisions, which was absent in the approach followed here. This again indicates a lapse that could have been avoided for greater interpretability, hence commanding better clinical trust and acceptance for deep learning models. Last but not least, the current study has focused on only individual architectures, such as CNN and VGG19, without the investigation of ensemble methods. Previous work by Le et al. (2020) has shown that ensemble models, which combined the predictions of multiple architectures, improved robustness and accuracy, especially for those hard-to-classify lesions. This lack of ensemble methods in this work is another future direction for investigation to further boost classification performance. In general, the contributions made in this work align with the literature at least when it comes to the efficiency of CNNs and preprocessing. However, there are a number of gaps in terms of the implementation of more advanced methodologies, including global average pooling, GAN-based augmentation, multimodal learning, and ensemble approaches. Further research in these aspects can build on the promising results achieved and further enhance the reliability, interpretability, and scalability of automated dermatologic diagnostic systems.

Conclusion

The research investigated addressing the challenges in skin lesion classification using CNN and transfer learning methods on the HAM10000 dataset. The dataset contained 10,015 dermatoscopic images representing seven lesion classes and was thus well-suited for training and evaluating models in dermatologic diagnostics. It further emphasized the importance of an automated system to overcome traditional methods of diagnosis, which were highly dependent on clinicians and were thus subjective and inconsistent.

The baseline CNN model performed pretty well, with a testing accuracy of 96.37%. This underlined the potency of a well-designed architecture in extracting and learning features from complex images. This architecture is supported at the back by convolutional and pooling layers tuned with care, which formed a very important part of capturing nuanced patterns in skin lesions. Preprocessing steps like random oversampling and data augmentation proved useful for dealing with class imbalance, enhancing model generalizability, especially in the case of rare lesion types such as dermatofibroma and vascular lesions.

The accuracy on the test set for transfer learning using VGG19 was 81.14%, which is very helpful in scenarios where computational resources are not abundant. This model, however, underperformed the baseline CNN, but quicker convergence and lower computational loads confirm the usefulness of pre-trained architectures in medical image analysis. These results confirm that transfer learning can be a complementary approach in dermatologic diagnostics, especially for applications with limited data or hardware.

The challenges were that even after preprocessing, the class imbalance was a problem and was resulting in misclassification of minority classes most of the time. That's where more advanced augmentation techniques such as GAN should become helpful to increase the count of underrepresented classes in an artificial manner. Additionally, generalization of models to clinical may not be possible since there may be differences in conditions and imaging modalities from what is represented in the dataset. This disparity underlines the importance of lightweight, adaptive architectures that can handle heterogeneous imaging scenarios.

Interpretability became another critical point, especially for clinical adoption. While the literature discussed Grad-CAM visualizations as tools to enhance transparency, they were not implemented in this study. The lack of tools for interpretability hints at future research that needs to bridge the gap between model performance and clinical trust.

Furthermore, the exclusion of metadata, including patient demographics and lesion localization, further limited the scope of multimodal learning, which has been clearly shown in previous studies to significantly improve diagnostic accuracy.

These findings are in line with the literature and further confirm the effectiveness of CNNs and transfer learning in dermatologic image classification while pointing to areas of further improvement. The research provides important lessons that can be applied to the design and optimization of deep learning models for skin lesion classification. However, it also puts into perspective the need for further advances in areas related to ensemble learning, multimodal integration, and model interpretability to obtain more robust, reliable, and scalable solutions for dermatologic diagnostics. This study further strengthens the transformative potential of AI in medical imaging, especially for early detection and diagnosis of skin cancer. Solving the identified challenges using emerging technologies will help take these findings a step further with future research, developing fully comprehensive, clinically viable instruments that enhance the accuracy, accessibility, and efficiency of dermatologic care.

References

https://www.kaggle.com/datasets/kmader/skin-cancer-mnist-ham10000

Adebiyi, A., Abdalnabi, N., Smith, E.H., Hirner, J., Simoes, E.J., Becevic, M. and Rao, P., 2024. Accurate Skin Lesion Classification Using Multimodal Learning on the HAM10000 Dataset. medRxiv, pp.2024-05.

Akter, M.S., Shahriar, H., Sneha, S. and Cuzzocrea, A., 2022, December. Multi-class skin cancer classification architecture based on deep convolutional neural network. In 2022 IEEE International Conference on Big Data (Big Data) (pp. 5404-5413). IEEE.

Al Mahmud, A., Azam, S., Khan, I.U., Montaha, S., Karim, A., Haque, A., Zahid Hasan, M., Brady, M., Biswas, R. and Jonkman, M., 2024. SkinNet-14: a deep learning framework for accurate skin cancer classification using low-resolution dermoscopy images with optimized training time. *Neural Computing and Applications*, *36*(30), pp.18935-18959.

Bansal, M., Kumar, M., Sachdeva, M. and Mittal, A., 2023. Transfer learning for image classification using VGG19: Caltech-101 image data set. *Journal of ambient intelligence and humanized computing*, pp.1-12.

Bechelli, S. and Delhommelle, J., 2022. Machine learning and deep learning algorithms for skin cancer classification from dermoscopic images. *Bioengineering*, *9*(3), p.97.

Faghihi, A., Fathollahi, M. and Rajabi, R., 2024. Diagnosis of skin cancer using VGG16 and VGG19 based transfer learning models. *Multimedia Tools and Applications*, *83*(19), pp.57495-57510.

Jain, S., Singhania, U., Tripathy, B., Nasr, E.A., Aboudaif, M.K. and Kamrani, A.K., 2021. Deep learning-based transfer learning for classification of skin cancer. Sensors, 21(23), p.8142.

Le, D.N., Le, H.X., Ngo, L.T. and Ngo, H.T., 2020. Transfer learning with class-weighted and focal loss function for automatic skin cancer classification. arXiv preprint arXiv:2009.05977.

Musthafa, M.M., TR, M., V, V.K. and Guluwadi, S., 2024. Enhanced skin cancer diagnosis using optimized CNN architecture and checkpoints for automated dermatological lesion classification. BMC Medical Imaging, 24(1), p.201.

Raghavendra, P.V., Charitha, C., Begum, K.G. and Prasath, V.B.S., 2023. Deep Learning–Based Skin Lesion Multi-class Classification with Global Average Pooling Improvement. Journal of Digital Imaging, 36(5), pp.2227-2248.

Sönmez, A.F., Çakar, S., Cerezci, F., Kotan, M., Delibaşoğlu, İ. and Çit, G., 2023. Deep Learning-Based Classification of Dermoscopic Images for Skin Lesions. *Sakarya University Journal of Computer and Information Sciences*, *6*(2), pp.114-122.

Tahir, M., Naeem, A., Malik, H., Tanveer, J., Naqvi, R.A. and Lee, S.W., 2023. DSCC_Net: multi-classification deep learning models for diagnosing of skin cancer using dermoscopic images. *Cancers*, *15*(7), p.2179.

Shetty, B., Fernandes, R., Rodrigues, A.P., Chengoden, R., Bhattacharya, S. and Lakshmanna, K., 2022. Skin lesion classification of dermoscopic images using machine learning and convolutional neural network. *Scientific Reports*, *12*(1), p.18134.

Tahir, M., Naeem, A., Malik, H., Tanveer, J., Naqvi, R.A. and Lee, S.W., 2023. DSCC_Net: multi-classification deep learning models for diagnosing of skin cancer using dermoscopic images. *Cancers*, *15*(7), p.2179.

Bechelli, S. and Delhommelle, J., 2022. Machine learning and deep learning algorithms for skin cancer classification from dermoscopic images. *Bioengineering*, *9*(3), p.97.

Sönmez, A.F., Çakar, S., Cerezci, F., Kotan, M., Delibaşoğlu, İ. and Çit, G., 2023. Deep Learning-Based Classification of Dermoscopic Images for Skin Lesions. *Sakarya University Journal of Computer and Information Sciences*, *6*(2), pp.114-122.

Debelee, T.G., 2023. Skin Lesion Classification and Detection Using Machine Learning Techniques: A Systematic Review. *Diagnostics*, *13*(19), p.3147.

Anand, V., Gupta, S., Koundal, D., Mahajan, S., Pandit, A.K. and Zaguia, A., 2022. Deep learning based automated diagnosis of skin diseases using dermoscopy. *Computers, Materials & Continua*, *71*(2), pp.3145-3160.

Adegun, A.A. and Viriri, S., 2020. FCN-based DenseNet framework for automated detection and classification of skin lesions in dermoscopy images. *IEEE Access*, *8*, pp.150377-150396.

Ahmed, I.A., Senan, E.M., Shatnawi, H.S.A., Alkhraisha, Z.M. and Al-Azzam, M.M.A., 2023. Multi-Models of Analyzing Dermoscopy Images for Early Detection of Multi-Class Skin Lesions Based on Fused Features. *Processes*, *11*(3), p.910.